Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq

Decoupling genetics, lineages and tumor micro-environment in gliomas by single-cell RNA-seq

Andrew S. Venteicher†[1,2,3], Itay Tirosh†*[2], Christine Hebert[1,2], Keren Yizhak[1,2], Cyril Neftel[1,2,4], Mariella G. Filbin[1,2,5], Volker Hovestadt[1,2], Leah E. Escalante[1,2], McKenzie L. Shaw[1,2], Christopher Rodman[2], Shawn M. Gillespie[1], Danielle Dionne[2], Christina C. Luo[1], Hiranmayi Ravichandran[1], Ravindra Mylvaganam[1], Christopher Mount[6], Maristela L. Onozato[1], Brian V. Nahed[3], Hiroaki Wakimoto[3], William T. Curry[3], A. John Iafrate[1], Miguel N. Rivera[1,2], Matthew P. Frosch[1], Todd R. Golub[2,5,7], Priscilla K. Brastianos[8], Gad Getz[1,2], Anoop P. Patel[3], Michelle Monje[6], Daniel P. Cahill[3], Orit Rozenblatt-Rosen[2], David N. Louis[1], Bradley E. Bernstein[1,2], Aviv Regev[1,7,9]‡*, Mario L. Suvà[1,2]‡*

[1] Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114, USA

[2] Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

[3] Department of Neurosurgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114 USA

[4] Institute of Pathology, CHUV, Faculty of Biology and Medicine, 1011 Lausanne, Switzerland

[5] Department of Pediatric Oncology, Dana-Farber Cancer Institute and Children's Hospital Cancer Center

Boston, MA 02215, USA

[6]Departments of Neurology, Neurosurgery, Pediatrics and Pathology, Stanford University School of Medicine, Stanford, CA, 94305, USA.

[7] Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA

[8] Departments of Medicine and Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, 02114 USA

[9] Koch Institute and Department of Biology, MIT, Cambridge, MA 02139, USA

†These authors contributed equally to this work.

‡These authors jointly supervised this work.

*Corresponding authors:

Email: Suva.Mario@mgh.harvard.edu (MLS); aregev@broadinstitute.org (AR); tirosh@broadinstitute.org (IT)

One Sentence Summary: single-cell RNA-seq redefines gliomas lineages

ABSTRACT

Tumor subclasses differ in the genotypes and phenotypes of malignant cells, and in the composition of the tumor microenvironment (TME). Here, we dissect these influences in IDH-mutant gliomas, combining 14,226 single-cell RNA-seq profiles from 16 patient samples with bulk RNA-seq profiles from 165 patient samples. Differences in bulk profiles between IDH-mutant astrocytoma and oligodendroglioma can be primarily

explained by distinct TME and signature genetic events, whereas both tumor types share similar developmental hierarchies and lineages of glial differentiation. As tumor grade increases, we find enhanced proliferation of malignant cells, larger pools of undifferentiated glioma cells and increase in macrophage over microglia programs in TME. Our work provides a unifying model for IDH-mutant gliomas and a general framework to dissect the differences between human tumor subclasses.

MAIN TEXT

Tumor fitness, evolution and resistance to therapy are governed by the combination of selection of cancer cells with specific genotypes, by expression programs related to cellular phenotypes and by influences of the tumor microenvironment (TME) (*1*). In recent years, studies such as those of The Cancer Genome Atlas (TCGA) have charted the genetic landscape and the bulk expression states of thousands of tumors, identifying driver mutations and defining tumor subtypes on the basis of specific transcriptional profiles (*2, 3*). While the genetic state of tumors could be studied with high precision, bulk expression profiles provide only limited insight as they average together the phenotypic determinants of cancer programs, TME influences and intra-tumoral genetic heterogeneity. Single-cell RNA-seq (scRNA-seq) can help address those challenges (*4-7*), but poses financial and logistic considerations, including the time required to accrue large cohorts of fresh tumor specimen for single cell analysis, especially in rare tumor types.

We reasoned that scRNA-seq of a limited number of representative tumors could be combined with existing bulk data from large cohorts to decipher these distinct effects, and sought to apply this approach to understand the differences between two major types of diffuse gliomas. In adults, diffuse gliomas are classified into three main categories on the basis of integrated genetic and histologic parameters: IDH-wildtype glioblastoma (GBM) is the most prevalent and aggressive form of the disease, while mutations in *IDH1/2* define two major classes of gliomas: astrocytoma (IDH-A) and oligodendroglioma (IDH-O) (*8*). IDH-A and IDH-O are two distinct tumor types that differ in their genetics, histopathology and prognosis. Genetically, IDH-A tumors are characterized by *TP53* and *ATRX* mutations, while IDH-O tumors are characterized by mutations in *TERT* promoter and loss of chromosome arms 1p and 19q, defining a robust genetic separation into two disease entities (*2*). In histopathology, IDH-A and IDH-O are distinct and thought to predominantly recapitulate astrocytic and oligodendrocytic lineages of glial differentiation, respectively. The notion that glial lineages differ between astrocytoma and oligodendroglioma, as implied by their names, originates from distinct morphology and tissue staining. However, expression of both oligodendroglial (*e.g.*, OLIG2) and astrocytic (*e.g.*, GFAP) markers can be readily identified in both diseases (*8*), mixtures of cells with histological features of neoplastic astrocytic and oligodendroglial cells are frequently observed within individual tumors, and cellular morphologies are only partially reminiscent of distinct glial cells, thus questioning the hypothesis of distinct glial lineages.

Here, we combine 9,879 scRNA-seq profiles from ten IDH-A with 4,347 single-cell profiles in six IDH-O and 165 TCGA bulk RNA profiles to decipher cancer cells genotypes and phenotypes and gain insight into TME composition across IDH-mutant gliomas. We find that differences in bulk profiles between IDH-A and IDH-O are primarily explained by signature genetic events and distinct TME composition, but not by distinct influences of glial lineages in the malignant cells of the two tumor types. Furthermore, as glioma grades increase, we observe both enhanced proliferation of malignant cells, a larger pool of undifferentiated glioma cells and an increase in macrophage over microglia programs in the TME. Our study redefines the cellular composition of human IDH-mutant gliomas and provides a general approach to decipher differences between tumor subtypes.

Deciphering differences between bulk IDH-mutant glioma samples with single-cell RNA-seq

We compared the expression profiles of IDH-A and IDH-O glioma using both bulk expression profiles from the TCGA datasets (76 IDH-O and 91 IDH-A gliomas) and newly measured single cell RNA-Seq profiles (**Fig. 1A**) from both tumor types. Comparing the TCGA bulk profiles, we found ~550 differentially expressed genes, suggesting distinct regulatory programs (*2*) (**Fig. 1B**). Since bulk profiles averaged the contributions of both genetically and phenotypically diverse malignant cells and additional diverse cells from the TME, we profiled single-cells from ten IDH-A tumors,

spanning clinical grades II-IV (**table S1, fig. S1**), retaining 6,341 single cell profiles after filtering out low-quality cells (**Fig. 1A**).

We first sought to classify single cells into malignant and non-malignant. While genetic mutations may be used for such classification, mutation calling from scRNA-seq has limited sensitivity and specificity - and combined single-cell DNA and RNA profiling is not yet scalable to thousands of cells (*9, 10*). We thus combined two complementary approaches. First, gene expression clustering separated cells into three groups, consistent with programs of glioma cells, immune cells and oligodendrocytes (**fig. S2**). Second, since glioma cells frequently harbor large-scale chromosomal aberrations (*2*), we estimated copy number variations (CNVs) from the average expression of genes in large chromosomal regions within each cell (*4*), and validated some of our predictions by whole exome sequencing and DNA FISH (**fig. S2**; **table S2)** (*11*). Expression-based and CNV-based classifications were highly consistent with one another, and we used both criteria to identify 5,097 malignant cells (**fig. S3**). Our classification scheme was further validated by IDH mutations whose detection, while technically limited in scRNA-seq data, was highly specific to cells classified as malignant (**fig. S3**; $P<10^{-16}$, hypergeometric test).

Many differences between bulk glioma samples do not stem from malignant cells

Surprisingly, when we directly compared the IDH-A malignant cells to 4,044 malignant cells profiled from six IDH-O tumors (*12*) (**Fig. 1B**), only approximately half of the genes

that were differentially expressed based on bulk TCGA samples were also differentially expressed between the single malignant cells of the two tumor types (**Fig. 1B, fig. S4**). This suggests that the remaining differentially expressed genes may reflect differences in the TME rather than differences in the expression programs of malignant cells. Indeed, most of the remaining expression differences between bulk samples involved either microglia/macrophage-specific genes or neuron-specific genes (*11*), which were preferentially expressed in bulk IDH-A or IDH-O samples, respectively (**Fig. 1C-E**, **fig. S4**), suggesting influences from non-malignant cells in the bulk profiles. Differential expression between IDH-A and IDH-O was consistent among microglia/macrophage-specific genes and among neuron-specific genes (**Fig. 1D**), allowing us to estimate the relative abundance of microglia/macrophages and of neuronal cells in each of the bulk tumors, from the average expression of these two signatures (**Fig. 1E**). Thus, IDH-A tumors are associated with more microglia/macrophages and less neuronal cells than IDH-O tumors, with few exceptions (**Fig. 1E**). Importantly, these differences are observed also between IDH-A and IDH-O tumors of the same clinical grade or when restricting the analysis to untreated tumors (**fig. S4**).

Genetic differences account for most of expression differences between malignant cells

Next, we focused on the expression differences between IDH-A and IDH-O that are significant both in comparison of bulk samples and of single malignant cells of the two

tumor types (*11*). We reasoned that specific genetic events might determine at least some of these differences. Indeed, most genes with higher expression in single malignant cells in IDH-A are located on chromosomes 1p and 19q, which are co-deleted in IDH-O (**Fig. 1F**). Loss-of-function of the transcriptional repressor *CIC*, which is specific to IDH-O, accounted for an additional ~10% of the expression differences (**Fig. 1F**), as inferred from a *CIC* expression signature (*11-13*). We also found a limited, yet significant, enrichment (P=0.018, hypergeometric test) of p53 targets among genes more highly expressed in IDH-O tumors, consistent with a mutated *TP53* in IDH-A. Overall, 57% of the expression differences were consistent with at least one of these genetic causes (**Fig. 1F**). Taken together, these results suggest that differences between bulk TCGA expression signatures of IDH-A and IDH-O primarily reflect TME composition and influences of genetic alterations.

scRNA-seq reveals shared glial lineages in IDH-A and IDH-O

IDH-A and IDH-O are thought to primarily recapitulate the astrocytic and oligodendrocytic glial lineages, respectively (*8*). However, the results above demonstrate that most differences between IDH-A and IDH-O may be accounted by genetics and TME, and question the hypothesis of distinct glial lineages in these tumors. Indeed, we observed limited differences in the expression of astrocyte-specific and oligodendrocyte-specific genes between IDH-A and IDH-O, either in bulk or in single cells profiles (**Fig. 2A**). Instead, the expression of these genes varied substantially across the cells within each of the IDH-A and IDH-O tumors. After

subtracting inter-tumor differences (*11*), principal component analysis (PCA) across all IDH-A cells demonstrated that PC1 and PC2 are associated with astrocyte-specific (PC1/2-high) and oligodendrocyte-specific (PC1/2-low) genes (**Fig. 2B**; **table S3**; $P<10^{-9}$, hypergeometric test). We refined the sets of glial lineage genes using the scRNA-seq data to define astrocyte-like and oligodendrocyte-like expression programs that co-vary across IDH-A cells (**Fig. 2C; table S3)** (*11*). These expression programs were neither accounted for by inter-tumor differences, nor by technical and batch effects (**fig. S5A,B and fig. S6A**), were reproduced in an analysis of 3,538 additional cells from two IDH-A tumors profiled with a different single cell RNA-seq protocol (**fig. S5C**), and were co-expressed also among IDH-O cells (**Fig. 2C**). We scored individual cells in each tumor type for expression of these programs, and classified cells with preferential expression of each program, as well as intermediate cellular states (**Fig. 2C**). All tumors exhibited a wide distribution of cellular states, yet there were more IDH-A cells in intermediate states (**Fig. 2C and fig. S6A**). Interestingly, the distribution of single cell profiles from IDH-wildtype GBMs differed, showing a bias towards the astrocytic program, supporting that the cellular architecture of IDH-A and IDH-O is specific to IDH-mutant tumors and is not shared across all diffuse gliomas (**fig. S6B**). Thus, our data supports a model in which malignant cells in IDH-A and IDH-O share inferred lineages of glial differentiation.

We next investigated whether the 192 genes differentially expressed between the malignant compartments of IDH-A and IDH-O (**Fig. 1F**) are shared across all malignant cells or whether they are specific to certain subpopulations. As expected, expression differences in 109 genes that can be attributed to signature genetic alterations (**Fig. 1F**)

were shared across all malignant cells (**fig. S6C**). However, differences between IDH-A and IDH-O in the expression of the remaining 83 differentially expressed genes (**table S3**) were most pronounced in differentiated tumor cells and almost completely abolished among the most undifferentiated cancer cells (**Fig. 2D**). Thus, undifferentiated cells from these tumor types exhibit increased similarity in gene expression programs, raising the possibility of shared cell-of-origin for IDH-A and IDH-O.

To further test this hypothesis, we analyzed DNA bulk methylation patterns, as DNA methylation may preserve epigenetic signatures of the cell-of-origin that are not evident by gene expression analysis. We found high similarity in DNA methylation between IDH-A and IDH-O compared to both IDH-wildtype gliomas and to IDH-mutant non-glioma tumors (**fig. S7**). While DNA methylation is highly influenced by the IDH mutation, this high similarity is consistent with a shared histogenesis of IDH-A and IDH-O.

Undifferentiated glioma cells are associated with proliferation and a shared stemness program

The high degree of expression similarity between undifferentiated cells in IDH-A and IDH-O and the possibility that these might reflect a stem/progenitor cell phenotype prompted us to further investigate their developmental programs. We previously identified cancer stem-like cells in IDH-O that display neural stem/progenitor programs and are highly enriched in cell cycle programs (*12*). Generalizing this finding across all IDH-mutant gliomas classes, we identified cycling cells on the basis of the expression of

consensus cell cycle signatures (**fig. S8A**) (*11, 12, 14*), and found that in both IDH-A and IDH-O only a small proportion of cells are proliferating (~4% on average in our cohort), and that there is an inverse correlation between proliferation and differentiation (**Fig. 3A**). Remarkably, the fraction of cycling cells for a given state of differentiation is similar between IDH-A and IDH-O (**Fig. 3A**). This supports a model in which proliferation and cell identity are tightly coupled in IDH-mutant tumors.

We derived a gene signature of the undifferentiated cells (excluding cycling cells) across the IDH-A and IDH-O tumors. Ninety genes were enriched within undifferentiated cells of at least three distinct tumors and were examined further for their co-expression among undifferentiated IDH-A and IDH-O cells (**Fig. 3B**). We defined a putative glioma stemness program as the subset of genes (**Fig. 3C**) that are both enriched and co-expressed in undifferentiated cells of both IDH-A and IDH-O. Indeed, this program includes neurodevelopmental transcription factors (*e.g.*, SOX4, SOX11 and TCF4), and is consistent with the expression program of human neural stem cells (NSCs) and neural progenitor cells (NPCs) and with a program we highlighted in IDH-O (**fig. S9**). We validated this tumor architecture in IDH-A tissues in fourteen additional cases (**table S1**), showing in each tumor: (i) two glial lineages of cancer cells differentiation, (ii) mutually exclusive expression of cycling (by Ki-67 staining) and differentiation (by ApoE expression) markers, and (iii) co-expression of cycling (Ki-67) and putative stem cell (SOX4) markers (**Fig. 3D, table S1**). This architecture has also been validated in a cohort of sixteen IDH-O (*12*).

Changes in tumor architecture associated with tumor grade and genetic subclones

While IDH-A and IDH-O share the same stem/progenitor programs and putative lineages of glial differentiation, our analyses reveal three inter-related differences: (**1**) the overall fraction of cycling cells (**Fig. S8**), and (**2**) of undifferentiated cells (**Fig. 2D**) are higher in our IDH-A cases; and (**3**) the two lineage scores are inversely related in IDH-O, consistent with a differentiation process in which one lineage represses the other, a relationship not observed in IDH-A (**fig. S6D-E**).

Notably, all three aspects vary significantly *within* the IDH-A tumors and correlate with tumor grade, such that higher grade tumors tend to have more cycling and undifferentiated cells and a more limited association between lineage programs (**Fig. 4A**, **fig. S10A-B**). This provides a molecular fingerprint for tumor progression, as IDH-A tumors begin as grade II lesions and progress to grade III and IV. We validated the correlation between the frequency of cycling malignant cells (as reflected by the cell cycle program) and tumor grade with analysis of bulk TCGA samples (**fig. S10C**).

We hypothesized that the observed fingerprint of tumor grade-associated changes might also be reflected in clonal evolution, whereby genetically distinct subclones within the same tumor vary in their frequency of cycling and undifferentiated cells, with selection favoring the more aggressive subclones. To study genetic *intra*-tumoral heterogeneity, we inferred CNVs with single cell expression profiles (**fig. S1B**), and predicted subclones in three of our tumors, MGH44, MGH57, and MGH103 (**Fig. 4B-C**,

fig. S2, S11). In each of these cases, while the overall tumor architecture was preserved across clones, we also observed variability either in the fraction of cycling cells or in differentiation patterns (**Fig. 4D-E**, **fig. S11)**. Overall, these cases together with two IDH-O cases (*12*), demonstrate that patterns of differentiation and proliferation can be partially modulated by genetics and be subjected to selection. Future studies should further investigate the modulation of our inferred cellular architecture by genetic evolution.

The microglia to macrophage balance in the glioma TME

Finally, we analyzed the diversity of microglia/macrophage cells, the predominant subset of non-malignant cells in the TME (n= 1,043 in IDH-A and 246 in IDH-O) using PCA (**fig. S12**). The second PC (PC2) reflected an inflammatory program consisting of cytokines (IL1, IL8, TNF), chemokines (CCL3, CCL4), NF_KB-related genes (REL, NFKBIA, NFKBIZ) and immediate early genes (JUNB, FOSB, EGR3, IER3, ATF3). The program was active in most microglia/macrophage cells across IDH-A and IDH-O tumors and is similar to a reported program in IDH-O (*12*) (**table S3**). PC1 highlighted two mutually opposing programs, which were highly consistent with microglia (PC1-high) and macrophage (PC1-low) expression programs (**Fig. 4A**, **table S3**). Top PC1-high genes included microglia markers, such as CX3CR1, P2RY12 and P2RY13 (*15*), whereas CD163, TGFBI and F13A1 were among the PC1-low genes and are more highly expressed in diverse macrophage populations than in microglia (*16*) (**Fig. 4A**).

Thus, PC1 may correspond to the differences between brain-resident microglia, and infiltrating macrophages that reach the tumor through the circulation and must pass through the blood-brain barrier.

However, scoring cells by the relative expression of microglia-specific to macrophage-specific genes revealed a continuum, rather than a bimodal distribution (**Fig. 4B**), which is difficult to account for by a simple model of two populations (microglia and macrophages) and suggests additional influences on these expression programs. Furthermore, even the top macrophage-like cells in gliomas have lower macrophage scores compared to macrophages from melanoma tumors (**Fig. 4C**) (*5*). Thus, the glioma microenvironment might have altered the expression profiles of macrophages, thereby decreasing their difference from microglia. Moreover, microglia/macrophages from each individual tumor had a limited range of scores, with some tumors biased towards macrophage-like cells (*e.g.*, MGH42) and others towards microglia-like cells (*e.g.*, MGH56) (**Fig. 4C**). This indicates that specific properties of the microenvironment of each tumor may be dominant over the immune cell-of-origin with respect to macrophage-like and microglia-like expression states, consistent with recent studies (*15*).

This observed inter-tumor variability in macrophage/microglia states correlated with grade, such that cells from higher-grade tumors were preferentially associated with macrophage-like expression states. We validated this association by comparing the

expression of macrophage-specific and microglia-specific genes across grades in bulk TCGA IDH-A and IDH-O tumors (**Fig. 4D**) and by RNA *in situ* hybridization (ISH) for CX3CR1 and CD163 in our own cohort (**Fig. 4F**). We also observed some cells that co-express microglia and macrophage programs in tumors, supporting our hypothesis of a continuum of microglia-like to macrophage-like states (**Fig. 4F**). These results suggest that early in their development, gliomas primarily contain brain-resident microglia-like cells, while macrophage-like programs are associated with higher grades, possibly coinciding with other grade-associated changes, such as increased angiogenesis and alterations of the blood brain barrier.

Accordingly, this effect may parallel changes in tumor vascularity. We derived a signature of endothelial-specific genes (*11*) and used their average expression to estimate the abundance of endothelial cells in each bulk tumor. This endothelial signature is correlated with the macrophage-specific, but not with microglia-specific, programs across IDH-O and IDH-A tumors (**Fig. 4E**). Moreover, the endothelial signature increases with tumor grade, paralleling changes in the macrophage-specific, but not microglia-specific, expression programs (**Fig. 4D**). While the endothelial program correlates with variability in the macrophage-like expression program between cells it does not account for the variability in the overall proportion of microglia and macrophages. IDH-A tumors have a considerably higher proportion of microglia+macrophage cells than IDH-O tumors, as noted above (**Fig. 1C**), and this difference is not accounted for by endothelial cells or by grade (**Fig. 4D**).

To search for additional mechanisms that might regulate infiltration of macrophage/microglia cells into the tumor we searched for genes that are not expressed by macrophage/microglia, but are correlated with the inferred abundance of macrophage/microglia cells across bulk tumor samples. We found 24 genes correlated with both microglia and macrophage expression across IDH-A tumors, and separately, across IDH-O tumors (**fig. S134A, top**). Although these analyses were performed within a tumor type and thus were not directly influenced by differences between IDH-A and IDH-O, these genes were preferentially expressed in IDH-A (**fig. S13A, bottom**), consistent with the increased macrophage/microglia signatures in IDH-A. While we cannot determine if these associations are causal (*i.e.*, we cannot distinguish whether these genes influence, or are influenced by, immune infiltration, or whether both are affected by a third hidden factor), the ability of this expression program to predict the extent of macrophage/microglia infiltration across tumors and tumor types (**fig. S13B**) suggests interactions between immune infiltration and other cells in the tumor. Interestingly, three of those genes were components of the complement system – a specialized arm of the innate immune system – which we recently observed in a similar analysis of fibroblast-immune cell interactions in melanoma (*5*).

Taken together, our observations (i) define microglia and macrophage programs in gliomas at single-cell resolution, (ii) associate the macrophage, but not the microglia program, with clinical grade and increased vascularity, (iii) highlight a continuity in transcriptional programs of microglia/macrophage in tumors (rather than a bimodal distribution), suggesting plasticity of cellular states, (iv) reveal an overall increase in

microglia/macrophage infiltration in IDH-A compared to IDH-O, and (v) define a tumor expression signature associated with increased microglia/macrophage infiltration.

Discussion

Our approach provides a general framework to decouple cancer cell genotypes, phenotypes, and the composition of the TME in tumors, combining single-cell analysis of a limited set of representative tumors with bulk samples collected for larger cohorts, such as those from TCGA. In IDH-mutant gliomas, our approach uncovers shared neural developmental programs and putative lineages of glial differentiation in IDH-A and IDH-O. Thus, IDH-mutant gliomas are primarily composed of three subpopulations of malignant cells including non-proliferating differentiated cells of two glial lineages, and proliferative undifferentiated cells that resemble neural stem/progenitor cells. The shared glial lineages and developmental hierarchies suggest a common progenitor for all IDH-mutant gliomas with NSC/NPC-like programs, shedding light on a long-standing debate in gliomagenesis (*17*).

Our study thus represents a shift in our understanding of the histogenesis of glial tumors and supports a model where IDH-mutant gliomas subclasses share developmental programs and putative lineages of glial differentiation, but differ primarily by genetic mutations and TME composition; all IDH-mutant gliomas we examined at single cell resolution, including 10 IDH-A and 6 IDH-O tumors, defined by genetics and histopathology, contained malignant cells recapitulating oligodendrocytic-like and

astrocytic-like glial programs as well as a neural precursor program. While our cohort is fairly limited, our cases have had little selection bias (consecutive cases operated at our institution), and our observations have been validated in larger cohorts by tissue staining and by analysis of the TCGA datasets.

Given the similar developmental architecture of IDH-A and IDH-O, the morphological differences between these two entities might be linked to genetic differences between IDH-A and IDH-O and to TME composition. Accordingly, at least two genes involved in cytoskeleton and cell shape are downregulated by IDH-O-specific mutations: (**I**) glial fibrillary acidic protein (GFAP), a marker commonly used to assess astrocytic lineage in histopathology, is positively regulated by CIC (*12*) and thus more highly expressed in IDH-A than IDH-O (**table S3**); and (**II**) RHOC, encoding RhoC GTPase, a well-known regulator of cell shape and motility (*18, 19*) is located on chromosome arm 1p and therefore more highly expressed in IDH-A (**table S3**). Thus, signature genetic events might influence the morphology of cancer cells and underlie at least some of the histopathologic differences.

We also found a considerable difference in the TME composition of IDH-mutant gliomas, whereby IDH-A is enriched with microglia/macrophages signatures. These differences in TME composition may also at least in part be driven by genetic influences. For example, TP53 (mutated only in IDH-A) has been implicated with effects on inflammation and immune infiltration (*20*).

While our data supports a shared architecture for all IDH-mutant gliomas, the cellular composition in other diffuse gliomas might differ. Indeed, we were not able to clearly identify a similar architecture in IDH-wildtype GBM. As much of the literature on putative glial lineages of malignant cells of gliomas preceded the discovery of the IDH1/2 mutations, IDH-wildtype GBM might have confounded analysis in those studies. By analyzing IDH-mutant gliomas of different clinical grades (spanning II-IV) at single cell resolution, we identified a potential molecular fingerprint of tumor progression, with support in TCGA datasets; our analyses suggest that high-grade lesions show increased proliferation, larger pools of undifferentiated cells, partially aberrant differentiation programs and increased infiltration by macrophages over resident microglia. Finally, from a therapeutic standpoint, our data raise the possibility that triggering cellular differentiation could arrest the growth of these tumors. By shedding light on the cellular composition of IDH-mutant gliomas, our data offer opportunities for the design of immunotherapies targeting cancer cells phenotypes, a potentially novel avenue in the treatment of these currently incurable malignancies.

Figure Legends

Figure 1. Expression differences between IDH-A and IDH-O are governed by the tumor microenvironment and genetics. (A) Workflow: Freshly-resected tumors were dissociated to single cell suspension, FACS-sorted and profiled by SmartSeq2 in 96-well plates. (**B**) Differential expression between IDH-A and IDH-O across bulk TCGA tumors (left), across single cells (middle) and the averages from each of these two analyses (right). (C) Differentially expressed genes by bulk analysis include microglia/macrophage-specific genes (left column) and neuron-specific genes (right column). (D) Distribution of expression differences between bulk IDH-A and IDH-O samples for microglia/macrophage-specific genes (black) and neuron-specific genes (grey). (E) Microglia/macrophage scores (X-axis) and neuron scores (Y-axis) (*11*), for bulk IDH-O (blue) and IDH-A (purple) tumors. (**F**) Left: Differentially expressed genes which are not microglia/macrophage-specific or neuron-specific assigned to four categories (top to bottom rows)of genetic influences (*11*): (**i**) genes residing in chromosome arms 1p or 19q, (**ii**) genes activated or (**iii**) repressed by CIC, and (*iv*) P53 target genes. Right: Observed and expected percentages of IDH-A specific genes assigned to the first two categories and IDH-O specific genes assigned to the last two categories. Expected percentages were defined by analysis of all genes rather than only the IDH-A and IDH-O specific genes.

Figure 2. Glial lineages are shared among IDH-A and IDH-O. (**A**) Average expression levels of oligodendrocytic-specific (light blue) and astrocytic-specific (black) genes across all IDH-A (Y axis) and IDH-O (X axis) malignant cells. (**B**) Correlations of oligodendrocytic-specific (light blue) and astrocytic-specific (black) genes with PC1 (X axis) and PC2 (Y axis) from a PCA of all IDH-A malignant cells. (**C**) Classification of malignant cells (columns) from IDH-A (left panel) and IDH-O (right panel), by the differential expression of 50 oligodendrocytic and 50 astrocytic genes. Bottom: relative expression of the 100 genes (rows); Top: significance of differential expression ($-\log_{10}$(P-value of a *t*-test)) between oligodendrocytic and astrocytic genes. Cells were sorted by significance from the most oligodendrocytic-like to the most astrocytic-like cells; dashed lines indicate a significance threshold of $P<0.01$. (**D**) For each malignant cell in IDH-A (purple) and IDH-O (blue), we present its differentiation scores (X-axis, maximum of oligodendrocytic and astrocytic scores) *vs*. the average expression of IDH-A (left) or IDH-O (right) specific genes (Y axis, excluding those genes exhibiting differential expression due to genetic alterations). Lines indicate the corresponding local weighted smoothing regression (LOWESS), demonstrating the decreased differences between IDH-A and IDH-O programs in cells with low glial differentiation scores.

Figure 3. Undifferentiated cells in IDH-A and IDH-O are associated with cycling cells and a putative stemness program. (**A**) Percentage of cycling cells (Y axis) in sliding windows of 200 cells ranked by differentiation scores (X axis) for either IDH-A (purple) or IDH-O (blue) malignant cells. (**B**) Pearson correlations (color bar) between

the expression profiles of ninety genes preferentially expressed in undifferentiated cells, across IDH-A (top) and IDH-O (bottom) undifferentiated cells. Genes are ordered by their correlation with the highest-scoring cluster in each analysis (*11*). (**C**) Pearson correlations of the ninety genes in (B) with the highest-scoring clusters in (B) in IDH-A (X-axis) and IDH-O (Y-axis). The top consistent genes are marked. (**D**) *In situ* RNA hybridization (ISH) shows mutually exclusive expression of astrocytic (APOE, blue) and oligodendrocytic (APOD, red) lineage markers; mutually exclusive expression of astrocytic and proliferation (Ki67, red, arrow) markers; and co-expression of proliferation and stem/progenitor (SOX4, blue, arrow) markers.

Figure 4. Analysis of tumor architecture by tumor grade and in genetic subclones. (A) The percentage of cycling cells (top), of undifferentiated cells (middle) and the negative correlation between the two lineage scores (bottom) are all associated with tumor grade ($P<0.05$, one-way ANOVA). For each feature, bars show the average value across groups of tumors defined by tumor type and grade. Error-bars indicate standard error. (B-C) CNV inference in MGH103 (B) and MGH57 (C) reveals large-scale CNVs which vary between cells of the same tumor. Cells were clustered based on their CNV patterns at specific chromosomal regions (black lines at top) to define putative subclones. (D,E) Comparison of the two lineage scores (left) and percentage of cycling cells (right) between the two subclones indicated for MGH103 (D) and for MGH57 (E). Significant differences are indicated (* - $P<0.05$, ** - $P<0.001$**; Kolmogorov-Smirnov test for lineages and hypergeometric test for cell cycle).

Figure 5. Microglia and macrophages across IDH-mutant gliomas. (**A**) Microglia (Y-axis) and macrophages (X-axis) expression levels (*21*) of genes with high (red) and low (blue) PC1-scores from PCA of tumor microglia/macrophages (**B**) Top: distribution of scores by average expression of microglia (PC1-high) *vs*. macrophage (PC1-low) genes (*11*). Bottom: differential expression of selected microglia- and macrophage-specific genes among all cells ranked by the scores at top. (**C**) Fraction (color code) of cells in bins of scores, as defined in (B, top) for each glioma (rows); Macrophages from melanoma (*5*) are included for reference (top row). Right: tumor grades. (**D**) Average endothelial scores (X-axis) vs. macrophage (left) or microglia (right) (Y-axis) across IDH-A (purple) and IDH-O (blue) tumors (gray, grade II; black, grade III; red, grade IV). Arrows indicate grade-specific changes associated with increased expression of endothelial program. (**E**) Correlation between endothelial scores and macrophage/microglia scores across all IDH-A (purple) or IDH-O (blue) bulk TCGA tumors. (**F**) *In situ* RNA hybridization for microglia (CX3CR1, blue) and macrophage (CD163, red) markers. Left panel: MGH56 contains predominantly microglia-like cells. Central panels: MGH43 contains microglia-like cells, macrophage-like cells (blood vessels, arrows) and cells expressing both markers (arrows). Right panel: MGH42 stains exclusively for CD163.

References and notes:

1. A. Kreso, J. E. Dick, *Cell stem cell* **14**, 275-291 (2014).
2. N. Cancer Genome Atlas Research *et al.*, *The New England journal of medicine* **372**, 2481-2498 (2015).
3. M. S. Lawrence *et al.*, in *Nature,* **505**, 495-501 (2014).
4. A. P. Patel *et al.*, in *Science,* **344**, 1396-1401 (2014).
5. I. Tirosh *et al.*, in *Science,* **352**, 189-196 (2016).
6. P. Dalerba *et al.*, in *Nature biotechnology,* **29**, 1120-1127 (2011).
7. D. A. Lawson *et al.*, in *Nature,* **526**, 131-135 (2015).
8. D. N. Louis, H. Ohgaki, O. D. Wiestler, W. K. Cavenee, *IARC: Lyon.*, (2016).
9. I. C. Macaulay *et al.*, in *Nat Methods,* **12**, 519-522 (2015).
10. I. C. Macaulay *et al.*, in *Nat Protoc,* **11**, 2081-2103 (2016).
11. Materials and methods are available as supplementary materials on *Science* Online.
12. I. Tirosh *et al.*, in *Nature,* **539**, 309-313 (2016).
13. V. Padul, S. Epari, A. Moiyadi, P. Shetty, N. V. Shirsat, *Genes, chromosomes & cancer* **54**, 725-733 (2015).
14. M. S. Kowalczyk *et al.*, in *Genome Res,* **25**, 1860-1872 (2015).
15. M. L. Bennett *et al.*, in *Proceedings of the National Academy of Sciences of the United States of America,* **113**, E1738-1746 (2016).
16. Y. Lavin *et al.*, *Cell* **159**, 1312-1326 (2014).
17. H. Zong, L. F. Parada, S. J. Baker, *Cold Spring Harbor perspectives in biology* **7**, (2015).
18. L. Sequeira, C. W. Dubyk, T. A. Riesenberger, C. R. Cooper, K. L. van Golen, *Clin Exp Metastasis* **25**, 569-579 (2008).
19. M. Tseliou *et al.*, in *Cell Physiol Biochem,* **38**, 94-109 (2016).
20. T. Cooks, C. C. Harris, M. Oren, in *Carcinogenesis,* **35**, 1680-1690 (2014).
21. O. Matcovitch-Natan *et al.*, *Science*, (2016).
22. S. Picelli *et al.*, in *Nat Protoc,* **9**, 171-181 (2014).
23. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, in *Genome biology,* **10**, R25 (2009).
24. B. Li, C. N. Dewey, in *BMC Bioinformatics,* **12**, 323 (2011).
25. N. Cancer Genome Atlas Research, *The New England journal of medicine* **368**, 2059-2074 (2013).
26. P. Guilhamon *et al.*, in *Nat Commun,* **4**, 2166 (2013).
27. M. J. Aryee *et al.*, in *Bioinformatics,* **30**, 1363-1369 (2014).
28. G. Mohapatra *et al.*, in *J Mol Diagn,* **8**, 268-276 (2006).
29. A. K. Shalek *et al.*, in *Nature,* **510**, 363-369 (2014).
30. Y. Zhang *et al.*, in *J Neurosci,* **34**, 11929-11947 (2014).
31. C. L. Wei *et al.*, in *Cell,* **124**, 207-219 (2006).
32. J. Ihmels *et al.*, in *Nature genetics,* **31**, 370-377 (2002).
33. L. P. J. van der Maaten, G. E. Hinton, *Journal of Machine Learning Research* **9**, 2579-2605 (2008).
34. J. Shin *et al.*, *Cell stem cell* **17**, 360-372 (2015).

Acknowledgments:

This work was supported by grants from the Smith Family Foundation (to M.L.S.), the Broad Institute Broad*next*10 program (to M.L.S. and O.R.R.), the National Brain Tumor Society (to M.L.S and D.N.L), the V Foundation for Cancer Research (to M.L.S.), the Merkin Institute Fellows at the Broad Institute of MIT and Harvard (to M.L.S.), the Rachel Molly Markoff Foundation (to M.L.S and B.E.B), the American Cancer Society (to M.L.S.), NIH-NCI brain cancer SPORE P50CA165962 (Developmental Research Project to M.L.S), and start-up funds from the MGH department of Pathology. A.S.V. was supported by the NIH R25 fellowship (NS065743) and research grants from the American Brain Tumor Association and Neurosurgery Research and Education Foundation. I.T. was supported by a Human Frontier Science Program fellowship and a Rothschild fellowship. C.N. was supported by the Placide Nicod Foundation. V.H. was supported by EMBO long-term fellowship. A.R. was supported by funds from the Howard Hughes Medical Institute, the Klarman Cell Observatory, STARR cancer consortium, NCI grant 1U24CA180922, NCI grant R33CA202820, by the Koch Institute Support (core) grant P30-CA14051 from the National Cancer Institute, the Ludwig Center and the Broad Institute. D.P.C was supported by Burroughs-Wellcome Fund CAMS (BWF 1007616.02). A.R. is a scientific advisory board member for ThermoFisher Scientific and Syros Pharmaceuticals and a consultant for Driver Group. Flow cytometry and sorting services were supported by shared instrumentation grant 1S10RR023440-01A1. M.M. was supported by the California Institute of Regenerative Medicine (CIRM)

grants RB4-06093 and RN3-06510 and the Virginia and D.K. Ludwig Fund for Cancer Research. Data generated for this study are available through the Gene Expression Omnibus (GEO) under accession number GSE89567.

SUPPLEMENTARY MATERIALS

Materials and Methods

Figs. S1 to S13

Tables S1 to S3